

Graphs Identified by Logics with Counting

Sandra Kiefer, Pascal Schweitzer and Erkal Selman

RWTH Aachen University

`{kiefer,schweitzer,selman}@informatik.rwth-aachen.de`

Ahornstraße 55, 52074 Aachen, Germany

March 31, 2015

Abstract

We classify graphs and, more generally, finite relational structures that are identified by C^2 , that is, two-variable first-order logic with counting. Using this classification, we show that it can be decided in almost linear time whether a structure is identified by C^2 . Our classification implies that for every graph identified by this logic, all vertex-colored versions of it are also identified. A similar statement is true for finite relational structures.

We provide constructions that solve the inversion problem for finite structures in linear time. This problem has previously been shown to be polynomial time solvable by Martin Otto. For graphs, we conclude that every C^2 -equivalence class contains a graph whose orbits are exactly the classes of the C^2 -partition of its vertex set and which has a single automorphism witnessing this fact.

For general k , we show that such statements are not true by providing examples of graphs of size linear in k which are identified by C^3 but for which the orbit partition is strictly finer than the C^k -partition. We also provide identified graphs which have vertex-colored versions that are not identified by C^k .

1 Introduction

The k -variable fragment of counting logic, denoted by C^k , is obtained from first-order logic by adding counting quantifiers but only allowing formulas that use at most k variables. These finite variable logics play a central role in the area of model-checking since for them the model-checking problem and the equivalence problem are solvable in polynomial time (see [11]). For a while, there was the hope that for some fixed k the logic C^k can distinguish every pair of non-isomorphic graphs. This would imply that the graph isomorphism problem is solvable in polynomial time. However, in 1992, it was shown by Cai, Fürer and Immerman [6] that $\Omega(n)$ variables are required to identify all graphs on n vertices. Since the examples presented in that paper consist of graph isomorphism instances which are actually known to be solvable in polynomial time, this also shows that C^k does not capture polynomial time.

Concerning C^k , there are striking connections to other seemingly unrelated areas. For example, there exist several Ehrenfeucht-Fraïssé type games characterizing C^k [6, 8, 13]. Also strongly related is the $(k - 1)$ -dimensional version of a well-known color refinement algorithm, named after Weisfeiler and Lehman by Babai (see [6]). It turns out that the $(k - 1)$ -dimensional Weisfeiler-Lehman algorithm does nothing else but partition $(k - 1)$ -tuples of vertices according to their C^k -types. Another surprising connection exists to linear programming. The k -th level of the Sherali-Adams hierarchy of a natural linear integer programming formulation of graph isomorphism essentially corresponds to the expressive power of C^k [2, 12].

Among the finite variable logics, the fragment C^2 has been of particular interest because it is known to be decidable [10], and the complexity of the decision problem has been studied extensively [22]. Numerous results for this logic are known, we refer the reader to a survey by Grädel and Otto [9]. In practice, due to its strength and the fact that it can be evaluated in almost linear time, the logic C^2 (more specifically, the corresponding 1-dimensional Weisfeiler-Lehman algorithm) is an essential subroutine in all competitive canonical labeling tools (see [20]). Very recent results concerning C^2 include a paper by Krebs and Verbitsky studying the quantifier depth of C^2 -formulas for C^2 -equivalence classes of graphs [16]. Kopczynski and Tan show that for every fixed C^2 -formula, the set of those n for which there is a structure with a universe of size n satisfying the formula is semilinear [15].

While the above results deal with the problem of distinguishing two graphs from each other using finite variable counting logics, in this paper we are concerned with the concept of distinguishing a graph from every other non-isomorphic graph. We say that the graph is **identified** by the logic. More formally, a graph (or a finite relational structure) G is identified by a logic L if there is a sentence φ in L such that every graph (or finite relational structure) which satisfies φ is isomorphic to G .

Of course every graph is identified by some first-order sentence. However, by [6], as mentioned above, there is no $k \in \mathbb{N}$ such that every graph is identified by some formula in C^k . Let us focus on the case $k = 2$. It is not difficult to see that all trees are identified by C^2 . Moreover, a graph is asymptotically almost surely identified by C^2 , that is, the fraction of graphs of size n which are not identified by C^2 tends to 0 as n tends to infinity [3]. Even more strongly, it is known [4] that the fraction of graphs which are not identified is exponentially small in n . Similarly, a regular graph is asymptotically almost surely identified by C^3 [17]. However, not all graphs are identified. (For C^2 , consider a cycle of length at least 6, for example.) The following question arises.

What is the structure of graphs that are identified by C^k ?

Our results. We study graphs that are identified by C^2 and provide a complete classification for them. This classification can be used to draw several conclusions about general properties of identified graphs. For example, one can derive that if an undirected graph is identified by C^2 , then the C^2 -partition classes of the vertices are exactly the orbits of the automorphism group of the graph. This corollary is neither true when considering finite (relational) structures nor when considering C^k with $k > 2$. For C^2 , we also conclude

that if an undirected graph is identified, every vertex-colored version of it is identified by C^2 as well. This statement holds for finite relational structures, too, but is again not true for C^k with $k > 2$. Using our classification, we show that in time $O((n + m) \log n)$ it is possible to determine whether an undirected graph is identified by C^2 .

To prove the correctness of our classification, we need explicit constructions that solve the inversion problem and the canonization problem for C^2 . The inversion problem asks whether to a certain invariant a graph (or more generally, a finite structure) can be constructed. In the case of C^2 , such an invariant is the count of the C^2 -types of pairs of vertices x, y . A celebrated result by Otto [21] shows that the inversion problem and the canonization problem for C^2 can be solved in polynomial time. As a by-product, our direct constructions provide an alternative proof for this. In fact, they show that the inversion problem for C^2 can be solved in linear time. Our constructions make use of circulant graphs and doubly-circulant graphs. With these, we observe that every C^2 -equivalence class contains a graph whose C^2 -partition classes are the orbits. More strongly, there is a single automorphism of the graph witnessing this. To achieve inversion for finite structures, we use an old 1-factorization construction due to Walecki (see [19]) that decomposes the complete graph K_{2n} into $2n - 1$ disjoint perfect matchings.

Building on the classification of graphs identified by C^2 , we also classify finite structures that are identified by C^2 . For graphs, there is only one special case that may appear within a C^2 -partition class (namely the cycle of length 5). However, for finite structures there are 7 different special cases, which are of sizes 3, 4, 5 and 6 for a C^2 -partition class. Our classification theorem describes how these may be combined to form structures that are identified by C^2 . Due to the nature of the different special cases, the classification is more involved (see Theorem 27). Nevertheless, we can show that one can decide in almost linear time whether a structure is identified by C^2 .

For the logics C^k with $k > 2$ we collect several negative results. One can first observe that the triangular graphs form an infinite non-trivial class of strongly regular graphs which are identified by C^3 , implying that any classification result would have to include non-trivial infinite families [7, 14].

Contrasting our results for C^2 , we provide examples of graphs that are identified by C^3 but for which conclusions analogous to the ones mentioned above do not hold. More specifically, we present graphs identified by C^3 for which even the logic C^k with k linear in the size of the graph does not correctly determine the orbit partition. This yields graphs which the logic C^k identifies, but for which not all vertex-colored versions are identified by C^k . These ideas are based on the construction by Cai, Fürer and Immerman [6].

The existence of these graphs highlights an important fact. Even if a graph is identified by the logic C^k , it is not clear that it is possible to take advantage of that in order to canonize the graph. This stands in contrast to a remark in [6] claiming that a graph G identified by C^k can be canonized in polynomial time. In fact, the crucial property required for the approach hinted at there to be successful is that all vertex-colored versions of G need to be identified by C^k . Indeed, if this property holds then a standard recursive individualization approach canonizes the graph G . It would suffice to show that for all vertex-colored versions of G the orbits are determined by the logic. However, with a slight

alteration of our construction, we obtain a graph G that is identified by C^k and whose orbits are correctly determined, but for some vertex-colored versions of G the orbits are not correctly determined.

Independently of our work, Arvind, Köbler, Rattan and Verbitsky [1] have investigated the structure of undirected graphs identified by C^2 obtaining results similar to the ones we provide in Section 4.

Organization of the paper. After providing preliminaries in Section 2, we give constructions for graphs and finite relational structures with given color degrees in Section 3, also solving the inversion problem. We then classify graphs identified by C^2 in Section 4 and use these results to classify identified finite relational structures in Section 5. Finally, in Section 6 we collect negative results on logics C^k for $k > 2$.

2 Preliminaries

Unless specified otherwise, by a **graph** we always mean a finite undirected graph without loops. We denote the vertex set of a graph G by $V(G)$ and the edge set by $E(G)$. For a subset P of $V(G)$, we denote by $G[P]$ the subgraph of G induced by P . The number of neighbors of a vertex is the **degree** of the vertex. If all vertices have degree k , we call the graph **k -regular**. A **(k, ℓ) -biregular graph G on bipartition (P, Q)** is a graph on vertex set $P \cup Q$ such that P and Q are independent sets, every vertex in P has exactly k neighbors in Q and every vertex in Q has exactly ℓ neighbors in P .

Two vertices $v, v' \in V(G)$ are in the same **orbit of G** (more precisely, in the same orbit of the automorphism group of G) if G has an automorphism φ such that $\varphi(v) = v'$. The **orbit partition** of G is the partition of $V(G)$ into the orbits of G .

A graph G is **identified** by a logic L if there is a sentence φ in L such that every graph which satisfies φ is isomorphic to G . A logic L **distinguishes** two graphs G and G' if there is a sentence φ in L such that $G \models \varphi$ and $G' \not\models \varphi$. We say that G and G' are **L -equivalent** if they are not distinguished by L .

The **k -variable counting logic**, denoted by C^k , is the k -variable fragment of first-order logic enriched by counting quantifiers. For every $t \in \mathbb{N}$ we have the counting quantifier $\exists^{\geq t}$. For a formula $\varphi(x)$ with the free variable x and for a graph G we have $G \models \exists^{\geq t} x \varphi(x)$ if and only if there are at least t vertices $v \in V(G)$ such that $G \models \varphi[v]$ (where, as usual, $\varphi[v]$ denotes the formula obtained by substituting v for the free variable x in φ). The (bound and free) variables in a C^k -formula are all from a fixed k -element set, say $\{x_1, \dots, x_k\}$, but they can be reused. For example, the formula

$$\exists x_1 \exists^{\geq 2} x_2 (E(x_1, x_2) \wedge \exists^{\geq 5} x_1 E(x_1, x_2))$$

is a valid C^2 -sentence which says that there is a vertex with at least 2 neighbors of degree at least 5. The **C^k -type** of a vertex v in G is the set of all C^k -formulas $\varphi(x)$ such that $G \models \varphi[v]$. The **C^k -coloring** of G is the coloring of each vertex with its C^k -type.

The C^k -**partition** of a graph G is the partition of its vertex set induced by their C^k -types. Similarly, the C^k -**type** of a tuple (v, w) is the set of C^k -formulas $\varphi(x, y)$ such that $G \models \varphi[v, w]$. For a vertex v or a pair of vertices v and w to obtain the **atomic C^k -type** we consider only quantifier-free C^k -formulas.

Let G be a graph and Π be a partition of $V(G)$. We say that Π is **equitable** if for all $P, Q \in \Pi$ and $v, v' \in P$, the vertices v and v' have the same number of neighbors in Q . A vertex coloring χ is called equitable if the partition induced by χ is equitable.

It is a well-known fact that the C^2 -partition of a graph is its **coarsest equitable partition**. The C^2 -partition of a graph with n vertices and m edges can be calculated in time $O((m+n) \log n)$ by the color refinement procedure (see [5]), also called naïve vertex classification or 1-dimensional Weisfeiler-Lehman algorithm.

2.1 Relational structures and partially oriented graphs

In order to simplify the task of generalizing our results to finite relational structures, we introduce edge-colored partially oriented graphs. For every structure \mathfrak{A} we define an edge-colored partially oriented graph $\text{ec-POG}(\mathfrak{A})$ that holds all the information about \mathfrak{A} expressible by using only two variables.

An **edge-colored partially oriented graph** (in short, an ec-POG) is an edge-colored directed graph (G, c) (with c an edge-coloring function) without loops such that for every $(v, w) \in E(G)$, it holds that if $(w, v) \in E(G)$ then $c((v, w)) = c((w, v))$. Slightly abusing terminology, we say that an edge $(v, w) \in E(G)$ is **undirected** if $(w, v) \in E(G)$ and **directed** otherwise. We accordingly draw (v, w) and (w, v) as one undirected edge between v and w (see Figure 1) and denote it by $\{v, w\}$. An ec-POG (G, c) is **complete** if for all $v, w \in V(G)$ with $v \neq w$ we have $(v, w) \in E(G)$ or $(w, v) \in E(G)$.

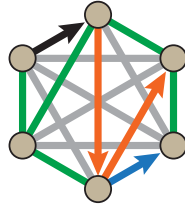


Figure 1: a complete ec-POG

In the following, we consider finite relational structures over a fixed signature $\sigma = (R_1, \dots, R_\ell)$ where R_i has arity r_i . The various definitions given for graphs are analogously defined for structures. Let \mathfrak{A} be a finite relational structure with universe A . For every $i \in \{1, \dots, \ell\}$ we define a function $c_i: A^2 \rightarrow \mathcal{P}(\{1, 2\}^{r_i})$ via

$$c_i(v_1, v_2) := \{(j_1, \dots, j_{r_i}) \in \{1, 2\}^{r_i} \mid \mathfrak{A} \models R_i(v_{j_1}, \dots, v_{j_{r_i}})\}$$

where, as usual, \mathcal{P} denotes the power set and $\{1, 2\}^{r_i}$ denotes the set of all r_i -tuples over $\{1, 2\}$. For every $v, w \in A$ with $v \neq w$ we let

$$c(v, w) := (c_1(v, w), \dots, c_\ell(v, w)).$$

Since for each i the possible images of c_i come from a set of bounded size, by using the order of the relations R_i in σ , one can easily define a canonical linear ordering \leq on the image of c (for example, by using the lexicographic order). With the help of this ordering, we define $\text{ec-POG}(\mathfrak{A}) := ((A, E_{\mathfrak{A}}), c_{\mathfrak{A}})$ as the complete ec-POG with vertex set A , edge set

$$E_{\mathfrak{A}} := \{(v, w) \mid v, w \in A, v \neq w \text{ and } c(v, w) \leq c(w, v)\}$$

and the edge coloring $c_{\mathfrak{A}} := c|_{E_{\mathfrak{A}}}$, the restriction of c to the domain $E_{\mathfrak{A}}$.

Note that $c(v, w)$ uniquely determines the atomic C^2 -types of v , w , (v, w) and (w, v) . Intuitively, this means that if the universe A has at least two elements, then the coloring c stores all the information about \mathfrak{A} that the logic C^2 can express. Hence, there is no need to include vertex colors or loops in our definition of ec-POGs. (More concretely, the atomic C^2 -type of v is encoded in the color $c(v, w)$ for all $w \neq v$, so does not need to be modeled by loops or vertex colors.) Note also that in $\text{ec-POG}(\mathfrak{A})$, a directed edge cannot have the same color as an undirected edge.

A partition Π of the vertex set of an ec-POG is **equitable** if for all $P, Q \in \Pi$, for all $v, v' \in P$ and for every edge color c , the vertices v and v' have the same number of c -colored outgoing, incoming and undirected edges connecting them to Q . An ec-POG is **color-regular** if for each edge color c every vertex v has the same c -indegree, the same c -outdegree and the same c -degree for undirected edges, respectively. An edge-colored undirected biregular graph on bipartition (P, Q) is called **color-biregular** if for every edge color c the subgraph induced by the edges of color c is biregular on (P, Q) . If the graph is partially oriented, vertices in each bipartition class must additionally have the same number of outgoing and incoming edges in each color. Note that for an equitable partition of an ec-POG, the graph induced by one C^2 -partition class is always color-regular and the graph induced between two C^2 -partition classes is color-biregular.

3 Inversion

In this section, we treat the so-called *inversion problem* that is closely related to the question which graphs are identified by C^2 . A **complete invariant** of an equivalence relation \equiv on a class \mathcal{C} of structures is a mapping \mathcal{I} from \mathcal{C} to some set S , such that $\mathfrak{A} \equiv \mathfrak{B}$ if and only if $\mathcal{I}(\mathfrak{A}) = \mathcal{I}(\mathfrak{B})$. We say that \mathcal{I} **admits linear time inversion** if given $s \in S$ one can construct in linear time a structure \mathfrak{A} with $\mathcal{I}(\mathfrak{A}) = s$ or decide that no such structure exists. This algorithmic task describes the inversion problem. Of course, if a structure \mathfrak{A} is identified then a solution to the inversion problem must construct \mathfrak{A} when given $\mathcal{I}(\mathfrak{A})$.

For C^2 , we show that a natural complete invariant, namely \mathcal{I}_C^2 , admits linear time inversion. Otto [21] proved that this invariant admits polynomial time inversion.

Given a graph G , one can define a linear ordering $P_1 \leq \dots \leq P_t$ on the classes of its coarsest equitable partition, which only depends on the C^2 -equivalence class of G (see [21]). This ordering allows us to define \mathcal{I}_C^2 , mapping G to (\bar{s}, M) , where \bar{s} is the

tuple $(|P_1|, \dots, |P_t|)$ and M is a $t \times t$ matrix, such that every vertex in P_i has exactly M_{ij} neighbors in P_j . It is easy to see that \mathcal{I}_C^2 is a complete invariant of C^2 .

3.1 Inversion for graphs

Definition 1. A graph is **circulant** if it has an automorphism with exactly one cycle. A graph on vertex set $P \dot{\cup} Q$ is **doubly-circulant** with respect to P and Q if it has an automorphism with exactly two cycles, one on P and the other on Q . Analogously, a graph is **multi-circulant** with respect to a partition $\{P_1, \dots, P_\ell\}$ of its vertex set if it has an automorphism with exactly ℓ cycles, each on one of the P_i .

It is well-known that circulant graphs can be constructed by numbering the vertices from 0 to $n - 1$, picking a set $S \subseteq \{1, \dots, \lfloor n/2 \rfloor\}$ of distances and inserting all edges between pairs of vertices whose distance of indices in the circular ordering is contained in S . By a double counting argument, one can easily see that for any k -regular graph on n vertices, the product $k \cdot n$ is even. Moreover, for all $k, n \in \mathbb{N}$ with $k < n$ and $k \cdot n$ being even, a circulant k -regular graph on n vertices can be constructed in linear time. Indeed, to obtain a k -regular graph in case k is even, one must only ensure that $|S| = k/2$. If k is odd (which implies that n is even), it suffices to have $n/2 \in S$ and $|S| = (k + 1)/2$ (see Figure 2a for $n = 12$ and $S = \{1, 2, 3, 6\}$). We call this the **circulant construction**.

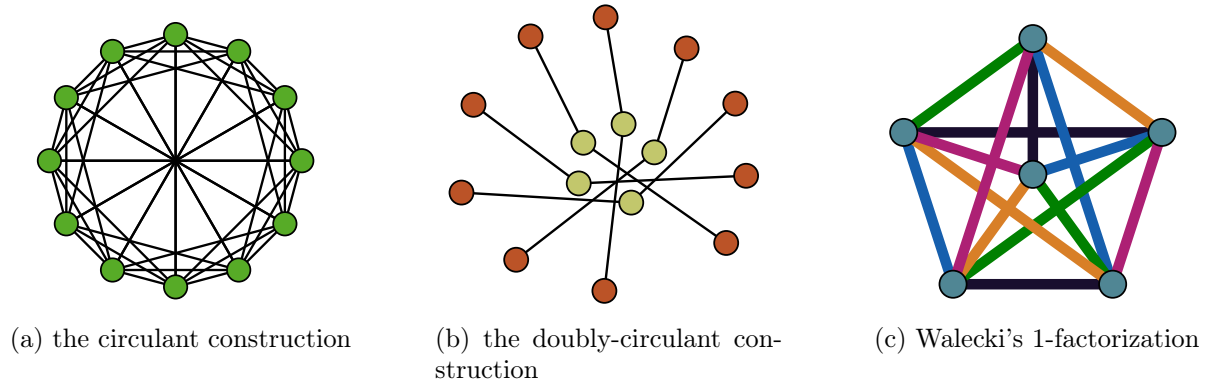


Figure 2

We have mentioned the necessary and sufficient evenness condition of $k \cdot n$ for a k -regular circulant graph on n vertices (with $k < n$). For the existence of doubly-circulant graphs, we also obtain a simple criterion. By double counting, every (k, ℓ) -biregular graph on a bipartition (P, Q) with $|P| = m$ and $|Q| = n$ satisfies $k \cdot m = \ell \cdot n$. The following lemma says that this condition is essentially sufficient to guarantee the existence of a doubly-circulant (k, ℓ) -biregular graph and that such a graph can be constructed in linear time.

Lemma 2. *For all $k, \ell, m, n \in \mathbb{N}$ with $k \leq n$, $\ell \leq m$ and $k \cdot m = \ell \cdot n$, one can construct in linear time a (k, ℓ) -biregular graph on a bipartition (P, Q) with $|P| = m$ and $|Q| = n$, which is doubly-circulant with respect to P and Q .*

Proof: Let $P := \{v_0, \dots, v_{m-1}\}$ and $Q := \{v'_0, \dots, v'_{n-1}\}$. We insert edges according to the following algorithm: Let j be the least integer such that the edge $\{v_0, v'_j\}$ has not yet been inserted. (So in the beginning, $j = 0$.) For every v_i , we insert edges to all vertices $v'_{(i+j+s) \bmod n}$, where s ranges over all multiples of m up to $\text{lcm}(m, n)$. Thus, a vertex of P in the obtained graph has degree $i_{P,Q} = \frac{\text{lcm}(m, n)}{m}$. Let $E_{P,Q}$ be the set of edges in the final graph we intend to construct. We have $|E_{P,Q}| = m \cdot k = n \cdot \ell$ edges, so $|E_{P,Q}|$ is a multiple of both m and n . Thus, $k = \frac{n \cdot \ell}{m} = \frac{|E_{P,Q}|}{m} \geq \frac{\text{lcm}(m, n)}{m} = i_{P,Q}$. If the inequality is proper, we can repeat the above process. In one iteration, the degree of a vertex in P increases by $\frac{\text{lcm}(m, n)}{m}$, a divisor of $\text{lcm}(m, n)$. Since $|E_{P,Q}|$ is a multiple of $\text{lcm}(m, n)$, we can therefore always ensure that the degree of a vertex in P never exceeds k . It strictly increases with every iteration, so after a finite number of iterations we obtain a graph H in which a vertex in P has exactly k neighbors in Q .

We provide an automorphism φ of H such that the orbits with respect to the cyclic group $\langle \varphi \rangle$ generated by φ are P and Q . Define φ such that it maps a vertex $v_i \in P$ to $v_{(i+1) \bmod m}$ and a vertex $v'_j \in Q$ to $v'_{(j+1) \bmod n}$. By construction, $\varphi(P) = P$ and $\varphi(Q) = Q$, so it remains to show that φ is a graph isomorphism, i.e., that $\{\varphi(u), \varphi(v)\} \in E_{P,Q}$ if and only if $\{u, v\} \in E_{P,Q}$ (this suffices since P and Q are independent sets). Let $v_i \in P$, $v'_j \in Q$. We have $\varphi(v_i) = v_{(i+1) \bmod m}$ and $\varphi(v'_j) = v'_{(j+1) \bmod n}$.

Suppose $\{v_i, v'_j\} \in E_{P,Q}$. Then $j = i + r + s \bmod n$ for some $r \in \{0, \dots, m-1\}$ and some multiple s of m with $s \leq \text{lcm}(m, n)$. This implies that $(j+1) = (i+1) + r + s \bmod n$ for the same r and s . Thus, $\{\varphi(v_i), \varphi(v'_j)\} \in E_{P,Q}$. By a symmetric argument, we get equivalence. \square

The construction described in the proof is depicted in Figure 2b.

The coarsest equitable partition of a graph is not necessarily its orbit partition. For example, on the disjoint union of two cycles of different sizes, the coarsest equitable partition is the unit partition, whereas the graph actually has two orbits. However, the above construction yields that for each C^2 -equivalence class, there is a representative whose coarsest equitable partition is indeed the orbit partition.

Theorem 3. *For every graph G , there is a C^2 -equivalent graph H which is multi-circulant with respect to its coarsest equitable partition.*

Proof: Let $\{P_1, \dots, P_t\}$ be the coarsest equitable partition of G . For every $i \in \{1, \dots, t\}$ the graph $G[P_i]$ is regular. By the circulant construction, there is a circulant graph $H(P_i)$ on each P_i , which is C^2 -equivalent to $G[P_i]$. By Lemma 2, we can connect $H(P_i)$ to $H(P_j)$ in such a way that the resulting graph on vertex set $P_i \dot{\cup} P_j$ is C^2 -equivalent to $G[P_i \cup P_j]$ and doubly-circulant with respect to P_i and P_j . In this way, by connecting the graphs $H(P_i)$ to each other, we get the desired graph H . \square

Corollary 4. \mathcal{I}_C^2 admits linear time inversion on the class of graphs.

Given an equivalence relation \equiv on a class \mathcal{C} of structures, the **canonization problem** for \equiv is the problem of finding a map $c: \mathcal{C} \rightarrow \mathcal{C}$ such that for every \mathfrak{A} in \mathcal{C} we have $c(\mathfrak{A}) \equiv \mathfrak{A}$ and for all $\mathfrak{A}, \mathfrak{B}$ in \mathcal{C} we have $c(\mathfrak{A}) \equiv c(\mathfrak{B})$. The map c is called a **canonization** for \equiv and $c(\mathfrak{A})$ is called the **canon** of \mathfrak{A} (with respect to c). Typically, the goal is to find such

a canonization c that can be evaluated efficiently.

As a by-product of the described constructions, we obtain the following corollary.

Corollary 5. *Canonization of graphs for C^2 can be done in time $O((n + m) \log n)$.*

Theorem 3 also immediately yields one of our main results.

Corollary 6. *If a graph G is identified by C^2 , then its coarsest equitable partition is the orbit partition.*

Note that the inverse of Corollary 6 does not hold: For example, on the graph consisting of two disjoint 3-cycles, the coarsest equitable partition is the orbit partition, but the graph is not identified by C^2 since it is C^2 -equivalent to a 6-cycle.

In Section 6, we also show that Corollary 6 does not hold for any logic C^k with $k \geq 3$.

Remark 7. It is natural to ask whether Corollary 6 holds for finite relational structures in general. This is not the case, not even for edge-colored undirected graphs. For example, let the colors of the edges of the complete graph K_6 correspond to a 1-factorization (see Figure 2c). There is only one 1-factorization of K_6 up to isomorphism (see [18]). This graph is identified by C^2 (see Theorem 24). It is rigid, i.e., its orbit partition is discrete, whereas its coarsest equitable partition is the unit partition.

3.2 Inversion for finite structures

Let $\mathfrak{A} = (A, R_1, \dots, R_\ell)$ be a finite relational structure. We define $\mathfrak{A}|_2$, the restriction of \mathfrak{A} to arity 2, to be the relational structure $(A, R'_1, \dots, R'_\ell)$ with

$$R'_i := \{(v_1, \dots, v_{r_i}) \in R_i \mid \{v_1, \dots, v_{r_i}\} \text{ has at most 2 elements,}\}$$

where r_i is the arity of R_i . Obviously, two relational structures \mathfrak{A} and \mathfrak{B} are C^2 -equivalent if and only if $\mathfrak{A}|_2$ and $\mathfrak{B}|_2$ are C^2 -equivalent. Furthermore, $\mathfrak{A}|_2$ and $\mathfrak{B}|_2$ are C^2 -equivalent if and only if $\text{ec-POG}(\mathfrak{A}|_2)$ and $\text{ec-POG}(\mathfrak{B}|_2)$ are C^2 -equivalent. Hence, the inversion problem for I_C^2 on finite relational structures reduces to the inversion problem for I_C^2 on ec-POGs.

The complete invariant I_C^2 that we defined on the class of graphs can naturally be extended to the class of ec-POGs: we simply replace the matrix in the original definition of I_C^2 with a matrix M such that for all $i, j \in \{1, \dots, k\}$ the entry M_{ij} is a tuple encoding for each edge-color d the number of d -colored outgoing edges from a vertex in P_i to P_j .

Similarly to the case of graphs, in order to solve the inversion problem for ec-POGs it suffices to solve it for the color-regular case and for the color-biregular case.

Color-regular case: Let n be the number of vertices. If n is odd, the degrees in each color in the underlying undirected graph must be even. Consequently, the circulant construction from Subsection 3.1 can be adapted to perform the inversion for directed and colored edges (see Figure 3). In more detail, for every map $\psi: \{1, \dots, n-1\} \rightarrow \{1, \dots, t\}$

such that for all i we have $\psi(n-i) = \psi(i)$ and for all j we have $|\psi^{-1}(j)| = d_j$, we obtain a circulant graph $\text{Circ}(\psi)$ by coloring the edge between vertices i and j satisfying $j \leq i$ with the color $\psi(i-j)$. This can also be interpreted as a 2-factorization of the graph into $(n-1)/2$ graphs that are 2-regular. The map ψ then dictates that the edges of the ℓ -th 2-factor are to be colored with color $\varphi(\ell)$.

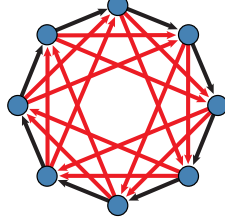


Figure 3: The circulant construction can be adapted to colored graphs as long as all color degrees are even.

If n is even and more than one color degree is odd, we cannot apply the circulant construction. In fact, as Remark 7 shows it might not be possible to construct a transitive graph with the given color degrees. Still, we can construct a canonical representative for the corresponding C^2 -equivalence-class as follows. We use the 1-factorization construction due to Walecki (see [19]). For even n the construction decomposes the complete graph K_n into $n-1$ disjoint perfect matchings. Geometrically, it is obtained as follows. Positioning the vertices $1, \dots, n-1$ in the plane to form a regular $(n-1)$ -gon and placing the vertex n into the center, each matching consists of an edge from n to some vertex $i \in \{1, \dots, n-1\}$ and all edges that are perpendicular to this edge (see Figure 2c). In algebraic terms, the edge $\{i, n\}$ is part of the i -th matching for $i \in \{1, \dots, n-1\}$. For $i, j \in \{1, \dots, n-1\}$ with $i \leq j$ either $|n-1-j+i|/2$ or $|i+j|/2$ is an integer ℓ and the edge $\{i, j\}$ is part of the ℓ -th matching. This shows that the factorization can be computed in linear time (see also [23, Example 7.1.2.]).

Suppose we want to have a graph with the specified color degrees d_1, \dots, d_t . By interpreting non-edges as edges of a special color, we can assume that $\sum_{i=1}^t d_i = n-1$. Then any function $\psi: \{1, \dots, n-1\} \rightarrow \{1, \dots, t\}$ with $|\psi^{-1}(j)| = d_j$ yields an edge-colored color-regular graph with the given color degrees if the edges of the i -th matching in the 1-factorization receive color $\psi(i)$. Note that we can choose ψ canonically by mapping the first d_1 integers to 1, the next d_2 integers to 2 and so on.

We thus obtain a canonical construction, which yields, for every set of given color degrees, an edge-colored color-regular undirected graphs. We call this the **matching construction** and denote for given color degrees d_1, \dots, d_t the canon obtained from an edge-coloring function ψ by $\text{Match}(\psi)$. The construction due to Walecki [19] has the property that two matchings of the 1-factorization always yield a Hamiltonian cycle. If we require directed edges, in which case in-degrees must be equal to out-degrees, we pair a suitable number of matchings and orient the obtained Hamiltonian cycles. With this construction it is directly possible to perform inversion of ec-POGs and thus of finite relational structures in linear time.

Color-biregular case: The doubly-circulant construction that was described in Subsection 3.1 for graphs can directly be altered to handle colored directed edges, by treating each color one after the other.

Corollary 8. \mathcal{I}_C^2 admits linear time inversion on the class of finite relational structures.

4 Characterization of the graphs identified by C^2

Here we examine the graphs that are identified by the logic C^2 and give a complete characterization of them. We derive various results from the characterization.

Definition 9. Let T be a tree with a designated vertex v . For $i \in \{1, \dots, 5\}$ let (T_i, v_i) be an isomorphic copy of (T, v) . Let F be the disjoint union of the five trees (T_i, v_i) and E be the edge set of a 5-cycle on vertex set $\{v_1, \dots, v_5\}$. Then we call $F + E$ a **bouquet**.

A **bouquet forest** is a disjoint union of vertex-colored trees and non-isomorphic vertex-colored bouquets.

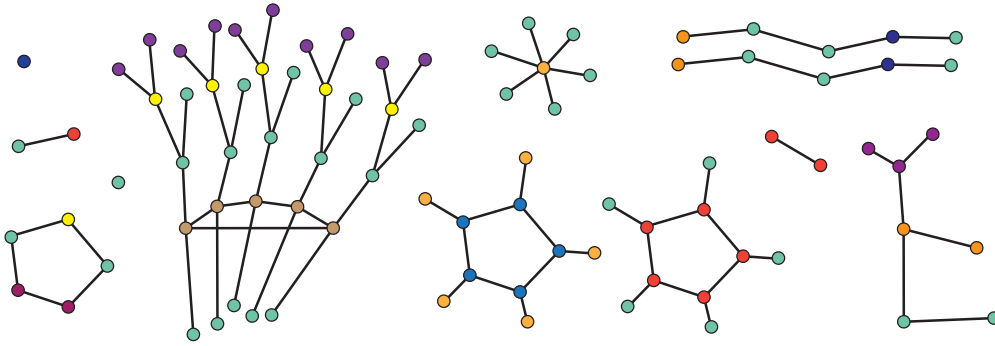


Figure 4: a bouquet forest with 4 bouquets

Notation 10. For sets P, Q we define $[P, Q] := \{\{v, w\} \mid v \neq w, v \in P, w \in Q\}$.

Definition 11. Let G be a graph with C^2 -coloring χ . We define the **flip of G** to be the vertex-colored graph (F, χ) with $V(F) = V(G)$ and $E(F) = E(G) \Delta ([P_1, Q_1] \cup \dots \cup [P_t, Q_t])$, where the (P_i, Q_i) are all pairs of (not necessarily distinct) C^2 -partition classes of G which satisfy

$$|[P_i, Q_i] \cap E(G)| > |[P_i, Q_i] \setminus E(G)|.$$

We say that (F, χ) is a **flipped graph**. If χ is the C^2 -coloring of F and (F, χ) is a flipped graph, we also say that F is flipped.

The symbol Δ in the definition denotes, as usual, the symmetric difference. To obtain the flip of G , we first calculate the C^2 -partition of G and check for each class P whether G has more edges in P than non-edges. If so, we take the complement of the edge relation there. We do the same for all pairs P, Q of distinct C^2 -partition classes, considering the edges in $[P, Q]$ there.

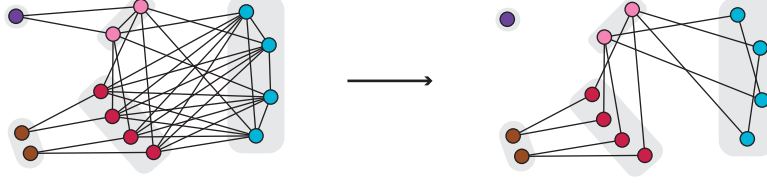


Figure 5: a graph with its C^2 -partition and its flip

While we may consider arbitrary vertex colorings in general, we want to stress the fact that the coloring χ of the flip (F, χ) of a graph G is always the C^2 -coloring of G . Note that χ induces an equitable partition of F , although it might not be the coarsest one. The aim of this section is to prove the following classification: a graph is identified by C^2 if and only if its flip is a bouquet forest. We first argue that we can restrict ourselves to flipped graphs.

Lemma 12. *A graph G is identified by C^2 if and only if its flip is identified by C^2 .*

Proof: Let G be a graph and χ its C^2 -coloring. Let (F, χ) be the flip of G . Note that $V(F) = V(G)$ by definition. Let G' be a graph on $V(G)$ with the same C^2 -coloring χ and let (F', χ) be the flip of G' . A map $f : V(G) \rightarrow V(G)$ is an isomorphism from (G, χ) to (G', χ) if and only if it is an isomorphism from (F, χ) to (F', χ) . This implies the contrapositives of both directions of the lemma. \square

For a graph G and its C^2 -coloring χ , we know that G and (G, χ) have the same automorphisms. If for a C^2 -partition class P of G , the induced subgraph $G[P]$ is not identified by C^2 , then G is not identified by C^2 . The same is true for subgraphs consisting of edges $[P, Q] \cap E(G)$ for distinct classes P and Q . We are going to exploit this simple fact in order to prove the classification. The following two lemmas provide strong restrictions to the edge relation of the regular and biregular graphs identified by C^2 .

Lemma 13. *A flipped regular graph is identified by C^2 if and only if it is a graph with no edges, a matching, or a 5-cycle.*

Proof: The backward direction is easy to check. For the forward direction, let G be a flipped k -regular graph on n vertices. Assume that G has at least one edge and it is neither a matching nor a 5-cycle. Then we have $n > 2$ and $k \geq 2$. In order to show that G is not identified by C^2 , we construct two non-isomorphic graphs H and H' , both C^2 -equivalent to G . Note that the graphs which are C^2 -equivalent to G are exactly the k -regular graphs on n vertices.

Since $k \geq 2$, we can apply the circulant construction from Subsection 3.1 in a way so that the resulting graph is connected. We choose H to be such a connected circulant graph.

For the construction of H' we do a case analysis:

1. k is odd: Then n is even.
 - (a) $n/2$ is even: By assumption, we have $k < n/2$. Let H' be the disjoint union of two k -regular graphs on $n/2$ vertices.
 - (b) $n/2$ is odd: Since k is odd and the graph is flipped, we have $2 < k < n/2 - 1$.

Let H' be the disjoint union of two k -regular graphs on $n/2 + 1$ and $n/2 - 1$ vertices, for example, circulant ones.

In both cases, we have $H \not\cong H'$ since H' is disconnected.

2. k is even:

- (a) n is even: Again we have $2 \leq k < n/2$. Since k is even, independent of the parity of $n/2$, there is a k -regular graph on $n/2$ vertices. We let H' be the disjoint union of two copies of such a graph. Again, we have $H \not\cong H'$ since H' is disconnected.
- (b) n is odd: Since G is flipped, we have $n \geq 2k + 1$.

If $n > 2k + 1$, from the parities of k and n we have $k < n/2 - 1$ and we can choose H' to be the disjoint union of a k -regular graph on $\lfloor n/2 \rfloor$ vertices and a k -regular graph on $\lceil n/2 \rceil$ vertices.

In case $n = 2k + 1$, there is no disconnected k -regular graph on $2k + 1$ vertices. Thus, in order to construct H' , take two disjoint k -cliques. Connect half of the first clique via a matching to half of the second clique. Lastly, connect all vertices in the unmatched halves to a new vertex v_0 .

Since G is not a 5-cycle by assumption, we cannot have $k = 2$. For $k > 2$, the vertex v_0 is the only vertex which does not belong to a k -clique. Hence, H' is not transitive, therefore not circulant. This implies $H \not\cong H'$.

□

Lemma 14. *The logic C^2 identifies a flipped (k, ℓ) -biregular graph if and only if $k \leq 1$ or $\ell \leq 1$. Furthermore, if the graph is not identified, there is a non-isomorphic graph, which is (k, ℓ) -biregular on the same bipartition (even if $k = \ell$).*

Proof: For the backward direction note that if $k \leq 1$ or $\ell \leq 1$, then G is a disjoint union of stars. Therefore, C^2 identifies G .

For the forward direction, let G be a flipped (k, ℓ) -biregular graph on bipartition (P, Q) . Without loss of generality assume that $k \geq \ell$. Suppose $k \geq 2$ and $\ell \geq 2$. In order to show that C^2 does not identify G , we construct non-isomorphic (k, ℓ) -biregular graphs H and H' , both on bipartition (P, Q) . We will define H and H' such that H is connected and H' is disconnected. Let $|P| = m$ and $|Q| = n$. Note that our assumption $k \geq \ell$ implies $n \leq m$.

Let H be a connected (k, ℓ) -biregular graph on bipartition (P, Q) . Such a graph exists due to the following argument. Starting with G , it suffices to explain how to repeatedly decrease the number of connected components while maintaining (k, ℓ) -biregularity. Since $k > 1$ and $\ell > 1$, each connected component has a cycle. Let C_1 and C_2 be two distinct connected components and let (p_1, q_1) and (p_2, q_2) be edges in these connected components, respectively, that lie on a cycle, where $p_i \in P$ and $q_i \in Q$. We replace the edges (p_1, q_1) and (p_2, q_2) by the edges (p_1, q_2) and (p_2, q_1) . This decreases the number of connected components by 1 and maintains vertex degrees. We conclude that there is a connected (k, ℓ) -biregular graph on bipartition (P, Q) .

We now define the graph H' . Suppose $P = \{v_0, \dots, v_{m-1}\}$ and $Q = \{w_0, \dots, w_{n-1}\}$. Let the complete bipartite graph on bipartition $(\{v_0, \dots, v_{\ell-1}\}, \{w_0, \dots, w_{k-1}\})$ be one of the connected components of H' . Since G is flipped, we have $k \leq n/2$ and $\ell \leq m/2$. It follows that $k \leq n - k$ and $\ell \leq m - \ell$. By $m \cdot k = n \cdot \ell$, we have $(m - \ell) \cdot k = (n - k) \cdot \ell$. By Lemma 2, there is a (k, ℓ) -biregular graph on bipartition $(\{v_\ell, \dots, v_{m-1}\}, \{w_k, \dots, w_{n-1}\})$. We take such a graph as the rest of H' . Obviously H' is disconnected. This completes the proof. \square

Notation 15. For a (k, ℓ) -biregular graph on bipartition (P, Q) we introduce the following notations.

$$\begin{aligned} P \square Q &: \iff k = \ell = 0 \\ P \dot{=} Q &: \iff k = \ell = 1 \\ P \ll Q &: \iff k \geq 2 \text{ and } \ell = 1 \end{aligned}$$

We also write $Q \gg P$ instead of $P \ll Q$.

With this notation, Lemma 14 says that a flipped graph is identified by C^2 if and only if for every two distinct C^2 -partition classes P and Q we have $P \square Q$, $P \dot{=} Q$ or $P \ll Q$.

For a graph G with C^2 -partition Π , we define the **skeleton** S_G of G as the graph with $V(S_G) = \Pi$ and $E(S_G) = \{\{P, Q\} \mid P \dot{=} Q \text{ or } P \ll Q \text{ in the flip of } G\}$.

Lemma 16. Let G be a flipped graph which is identified by C^2 . Then the following hold:

1. there is no path P_0, P_1, \dots, P_t in S_G with $P_0 \ll P_1$ and $P_{t-1} \gg P_t$,
2. there is no path P_0, P_1, \dots, P_t in S_G where $P_0 \ll P_1$ and P_t induces a 5-cycle or a matching,
3. in every connected component of S_G there is at most one vertex P that induces a matching or a 5-cycle in G .

Proof: For each condition, assuming it does not hold, we will define non-isomorphic graphs H and H' , both C^2 -equivalent to a subgraph of G that is induced by some classes of the C^2 -partition of G .

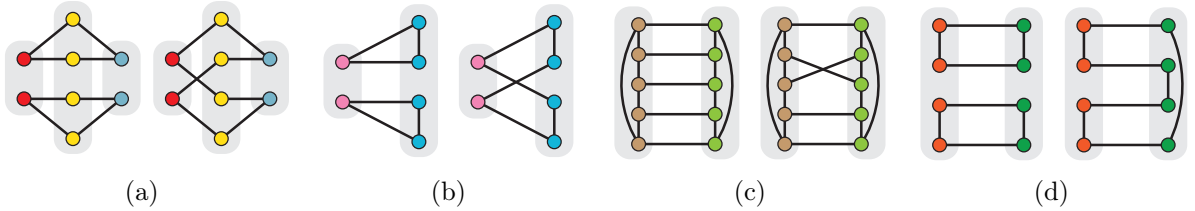


Figure 6: pairs of non-isomorphic graphs illustrating the necessity of the conditions in Lemma 16

Notation. For two C^2 -partition classes P and Q , let i_{PQ} denote the number of neighbors that a vertex in P has in Q .

(Part 1) First assume that $t = 2$ and $G[P_1]$ has no edges. Assume that in G we have $P_0 \ll P_1 \gg P_2$. We may assume $|P_0| \leq |P_2|$. Let $p := i_{P_0 P_1}$ and $q := i_{P_2 P_1}$. Then $p \geq q$.

In the following, Π , Θ and their indexed variants will always denote partitions of P_1 into classes of size p and q , respectively. Furthermore, we denote by $H(\Pi, \Theta)$ the graph on vertex set $P_0 \cup P_1 \cup P_2$ with the following properties.

- For every $i \in \{0, 1, 2\}$ the subgraph induced by P_i has no edges,
- we have $P_0 \ll P_1 \gg P_2$,
- the classes of Π form the neighborhoods in P_1 of vertices in P_0 ,
- the classes of Θ form the neighborhoods in P_1 of vertices in P_2 .

First we define H . Because $p \geq q$, there are partitions Π_0 and Θ_0 of P_1 such that there are at most $|P_0| - 1$ classes of Θ_0 that intersect two or more elements of Π_0 . One way to construct them is to number the vertices in P_1 and then to define the i -th element of Π_0 (and Θ_0) as the set of the i -th p (and q) vertices. We let $H := H(\Pi_0, \Theta_0)$.

We now define H' . We know that $|P_0| \geq 2$ since G is flipped. Hence, since $|P_0| \cdot p = |P_1| \cdot q$, there are partitions Π_1 and Θ_1 such that all classes of Θ_1 intersect two or more elements of Π_1 . We let $H' := H(\Pi_1, \Theta_1)$. Since $|P_2| > |P_0| - 1$, this implies $H \not\cong H'$ (see Figure 6a). This construction can easily be generalized for the case $t > 2$.

(Part 2) Without loss of generality, we may assume that $t = 1$. First let $G[P_1]$ be a 5-cycle. Since $P_0 \ll P_1$ and 5 is prime, the set P_0 is a singleton. This contradicts the assumption that G is flipped.

Let now $G[P_1]$ be a matching. We let H be the graph $G[P_1] + B$ on vertex set $P_0 \cup P_1$, where B is a subset of $[P_0, P_1]$ such that we have $P_0 \ll P_1$ in H and for each $v \in P_0$ the number of edges in $G[P_1]$ that v is connected to via B is as small as possible. (If $i_{P_0 P_1}$ is even, this number is $i_{P_0 P_1}/2$, otherwise it is $(i_{P_0 P_1} + 1)/2$.)

We let H' be the graph $G[P_1] + B'$, where B' is a subset of $[P_0, P_1]$ such that we have $P_0 \ll P_1$ in H' and there is a vertex in P_0 which is connected to $i_{P_0 P_1}$ edges in $G[P_1]$ via B' . Such a B' exists because P_0 cannot be a singleton. Figure 6b shows H and H' for $t = 1$ and $|P| = 4$.

(Part 3) Suppose P_0, P_1, \dots, P_t is a path in S_G such that both P_0 and P_t induce a matching or a 5-cycle in G . By Part 2, we have $P_0 \doteq P_1 \doteq \dots \doteq P_t$ in G . By Lemma 13, without loss of generality, we may assume that for every $1 \leq i \leq t - 1$, the edge set of $G[P_i]$ is empty. First, let $G[P_0]$ be a 5-cycle. Then P_t also consists of 5 elements and cannot be a matching. Assume that $G[P_t]$ is a 5-cycle. Let \tilde{H} be a graph on vertex set $P_0 \dot{\cup} \dots \dot{\cup} P_t$, such that for every $i \in \{1, \dots, t\}$ we have $P_{i-1} \doteq P_i$ in \tilde{H} and the graph $\tilde{H}[P_i]$ has no edge. Let C be a 5-cycle on P_0 . Then there is a unique cycle C' on P_t such that whenever $\{v, v'\}$ is an edge in C and $w, w' \in P_t$ are the two unique vertices connected via paths in \tilde{H} to v and v' , respectively, then $\{w, w'\} \in C'$. Let C'' be a 5-cycle on P_t that is different from C' . We define $H := \tilde{H} + C + C'$ and $H' := \tilde{H} + C + C''$. Then H and H' are non-isomorphic (see Figure 6c for $t = 1$).

Assume now that $G[P_0]$ is a matching. Note that $|P_0|$ must be even. Obviously, $G[P_t]$

cannot be a 5-cycle, so assume it is a matching. Let M be a matching on P_0 and let \tilde{H} be defined as above. Then there is a unique matching M' on P_t such that whenever $\{v, v'\}$ is an edge in M and $w, w' \in P_t$ are the two unique vertices connected via edges in \tilde{H} to v and v' , respectively, then $\{w, w'\} \in M'$. Let M'' be a matching on P_t that is different from M' . Because G is flipped, $|P_t| = 2$ is impossible. Hence, such a matching M'' exists. We define $H := \tilde{H} + M + M'$ and $H' := \tilde{H} + M + M''$ (see Figure 6d for $t = 1$ and $|P_0| = 4$). \square

Now we have all the ingredients to prove the main theorem of this section.

Theorem 17. *A graph is identified by C^2 if and only if its flip is a bouquet forest.*

Proof: We first show the backward direction. By Lemma 12, it suffices to show that every bouquet forest is identified by C^2 . This can be done by induction. The base case is that disjoint unions of non-isomorphic vertex-colored 5-cycles are identified by C^2 . Note that, in the definition of a bouquet forest, we do not allow isomorphic vertex-colored bouquets. (The reason for this is that C^2 cannot distinguish two 5-cycles from a 10-cycle.) For the inductive step, it suffices to observe that a graph is identified by C^2 if all vertex-colored versions of the graph obtained by removing all leaves of the graph are identified.

For the forward direction, let G be a graph identified by C^2 . Let (F, χ) be the flip of G . By Lemma 12, we know that (F, χ) is identified by C^2 .

By Lemma 13, for every color class P of (F, χ) the graph $F[P]$ either has no edges or is a matching or a 5-cycle.

By Lemma 14, for every two distinct color classes P and Q of (F, χ) we have $P \sqcap Q$, $P \doteq Q$ or $P \ll Q$.

We now show that the skeleton S_G is a forest.

First let P_0, P_1, \dots, P_t with $t \geq 2$ be color classes of (F, χ) with $P_0 \doteq P_1 \doteq \dots \doteq P_t$. We claim that $P_0 \doteq P_t$ is impossible. Assume $P_0 \doteq P_t$ for the sake of a contradiction. Note that $|P_0| \neq 1$ because F is flipped. So we have $|P_0| \geq 2$. Then we can arrange the edges in the given matchings as $|P_0|$ disjoint cycles each of size $t + 1$, but also as a single cycle of size $(t + 1) \cdot |P_0|$. This implies that (F, χ) is not identified by C^2 . By Lemma 12, we can conclude that G is not identified by C^2 , a contradiction.

Let P_0, \dots, P_t be a sequence of color classes such that for every $i \in \{1, \dots, t\}$ we have $P_{i-1} \ll P_i$ or $P_{i-1} \doteq P_i$. Assume that $P_{i-1} \ll P_i$ for at least one $i \in \{1, \dots, t\}$. Then by cardinality reasons it cannot be the case that $P_t \doteq P_0$ or $P_t \ll P_0$.

Also note that if $P_0 \ll \dots \ll P_t$ with $t \geq 2$, then Part 1 of Lemma 16 implies that $P_0 \ll P_t$ is impossible.

We conclude that S_G is a forest.

Furthermore, by Part 3 of Lemma 16, for every connected component T of S_G there can be at most one class $P \in V(T)$ such that $F[P]$ is a matching or a 5-cycle. To see that each connected component that contains a 5-cycle is a bouquet, note that by Corollary 6, all vertices of the 5-cycle must be in the same orbit and thus, the five trees obtained by

deleting the 5-cycle must be isomorphic. With Part 2 of Lemma 16, it follows that the subgraph of F induced by the vertices of T is either a forest or a bouquet. Moreover, no C^2 -partition class of F contains more than one 5-cycle, since this would contradict the fact that (F, χ) is identified. Hence, different 5-cycles must have different colors. This implies that there are no isomorphic colored bouquets in (F, χ) . \square

Corollary 18. *Given a graph with n vertices and m edges, we can decide whether it is identified by C^2 in time $O((m + n) \log n)$.*

Proof: Color refinement can be performed in time $O((m + n) \log n)$ (see [5]). Flipping a graph and checking whether the flip is a bouquet forest can be done in linear time. \square

A second corollary of Theorem 17 is concerned with vertex colorings of graphs that are identified by C^2 .

Corollary 19. *Let (G, χ) be a vertex-colored graph which is identified by C^2 and let χ' be a vertex coloring of G which induces a finer partition on $V(G)$ than χ does. Then (G, χ') is also identified by C^2 .*

Proof: Note that the statement is true if G is a bouquet forest. For the general case, we may assume that χ and χ' are equitable. Let (F, χ) be the flip of G and let (F', χ') be the flip of (G, χ') . Then the flip of (F, χ') is equal to the flip of (F', χ') . Since F is a bouquet forest, we conclude that (F, χ') is identified and thus, by applying Lemma 12 twice, we have that (F', χ') is identified. Again by Lemma 12, the vertex-colored graph (G, χ') is identified. \square

In the corollary, the vertex colorings χ and χ' of the graph G can be arbitrary, in particular χ can be monochromatic. We show in Section 6 that even when χ is required to be monochromatic this result cannot be generalized to the logics C^k for $k > 2$.

Since they can appear only once per connected component of the skeleton, we call a C^2 -partition class that is a 5-cycle or a matching an **exception**. Our classification of finite relational structures that are identified by C^2 , which we give in the next section, depends on the structural properties proven about identified graphs in this section. We summarize them for convenience.

Corollary 20. *A flipped graph G is identified by C^2 if and only if the following hold:*

1. *Each C^2 -partition class induces a graph identified by C^2 (i.e., the induced graph has no edges or it is a matching or a 5-cycle),*
2. *for all C^2 -partition classes P and Q we have $P \sqsubseteq Q$, $P \doteq Q$, $P \ll Q$ or $Q \ll P$,*
3. *the skeleton S_G is a forest,*
4. *there is no path P_0, P_1, \dots, P_t in S_G with $P_0 \ll P_1$ and $P_{t-1} \gg P_t$ in G ,*
5. *there is no path P_0, P_1, \dots, P_t in S_G where $P_0 \ll P_1$ and $G[P_t]$ is a 5-cycle or a matching, and*
6. *in every connected component of S_G there is at most one exception (i.e., a class P that induces a matching or a 5-cycle).*

5 General finite structures

Generalizing our observations about graphs, we now classify which finite relational structures are identified by C^2 .

Outline of the classification. To achieve our classification we proceed as follows. As previously observed (see Subsection 2.1), it suffices to analyze which edge-colored partially oriented graphs (i.e., ec-POGs) are identified. We would essentially like to follow the outline of the proof of the classification for graphs presented in Section 4. That is, we first intend to characterize which color-regular ec-POGs are identified, and then which color-biregular ec-POGs are identified by C^2 . Note that color-regular ec-POGs are exactly the ec-POGs which have only one C^2 -partition class. As we argue later, for an ec-POG that is identified, the underlying undirected graph is also identified. Exploiting this observation we will first consider only undirected graphs. In our proof for the color-regular case, we need the undirected color-biregular case, so we treat it first. We then generalize the results to color-regular ec-POGs and finally assemble these observations to classify all ec-POGs that are identified by C^2 .

Throughout this section we consider only complete ec-POGs. We can do this without loss of generality since we can interpret non-edges as edges of a particular color. The proofs of this section depend on the results from Section 4. We employ these results by interpreting uncolored graphs with edges and non-edges as complete graph with two edge colors.

In agreement with this outline we now start by analyzing color-biregular ec-POGs that do not have directed edges, i.e., undirected edge-colored color-biregular graphs. Recall that $K_{3,3}$ is the complete bipartite graph with partition classes of size 3.

Lemma 21. *Let G be an undirected complete bipartite edge-colored color-biregular graph on bipartition (P, Q) . Suppose G has at least three edge colors. Then there is a bipartite graph G' on bipartition (P, Q) which is non-isomorphic and C^2 -equivalent to G if and only if G is not isomorphic to an edge-colored $K_{3,3}$.*

Proof: It is straightforward to check that for every regular edge coloring of $K_{3,3}$ there is no C^2 -equivalent bipartite non-isomorphic graph. We thus focus on the converse.

Let G be undirected complete edge-colored color-biregular on bipartition (P, Q) . Suppose $|P| = p \cdot k$ and $|Q| = q \cdot k$ with co-prime p and q .

Let $1 \leq d_1^P \leq d_2^P \leq \dots \leq d_t^P$ and $1 \leq d_1^Q \leq d_2^Q \leq \dots \leq d_t^Q$ be the color degrees of colors $1, \dots, t$ for vertices in P and Q , respectively. Since $t \geq 3$, it holds that $|P| \geq 3$ and $|Q| \geq 3$. By double counting, we have that q divides d_i^P and p divides d_i^Q for all $i \in \{1, \dots, t\}$.

(Case $|P| = |Q|$.) We first consider the case $|P| = |Q|$. Note the following observation. If we construct a color-regular graph on (P, Q) for some subset of the colors with the correct color degrees, then it is always possible to complete the graph such that it has the correct color degrees for all edge colors. The reason is König's Theorem, which implies that a regular bipartite graph contains a perfect matching.

Assume $d_1^P = d_2^P = 1$ (which implies $d_1^Q = d_2^Q = 1$). If $|P| = 3$, the graph is an edge-colored $K_{3,3}$. However, if $|P| > 3$, there are two non-isomorphic (not necessarily complete) bipartite graphs with edge colors 1 and 2 and color degrees d_1^P and d_2^P : we can form a color-alternating Hamiltonian cycle or two shorter color-alternating cycles. By König's Theorem, these graphs can be extended and we conclude that for a bipartite graph with $|P| = |Q| > 3$ and $d_1^P = d_2^P = 1$ there is a C^2 -equivalent non-isomorphic graph on the same bipartition. Similarly, by Lemmas 12 and 14, if $1 < d_1^P < |Q| - 1$, then there are two non-isomorphic (d_1^P, d_1^Q) -biregular graphs on bipartition (P, Q) which can both be extended. Note that since there are at least three colors, $d_1^P < |Q| - 1$, so this resolves the case $|P| = |Q|$.

(Case $|P| \neq |Q|$.) Suppose now that $|P| \neq |Q|$. Consider a bipartite graph G' on (P', Q') with $|P'| = |Q'| = |P|/p = |Q|/q = k$ and with color degrees $d_1^P/q, \dots, d_t^P/q$ and $d_1^Q/p, \dots, d_t^Q/p$. From G' we obtain a graph with bipartition classes of sizes $|P|$ and $|Q|$ by replacing every vertex v in P' by p copies that each have in every color the same neighbors as v and then replacing each vertex w in Q' by q copies that each have in every color the same neighbors as w .

If we perform the construction twice with non-isomorphic graphs G'_1 and G'_2 then we obtain two non-isomorphic C^2 -equivalent graphs G_1 and G_2 . We already know from the first part of the proof that such non-isomorphic graphs G'_1 and G'_2 exist if $k \neq 3$.

It remains to consider the case $k = 3$. Since we have already treated $K_{3,3}$, we can assume without loss of generality that $p > 1$. As above, we will construct two non-isomorphic graphs G'_1 and G'_2 where the second bipartition class Q' has size $|Q'| := |Q|/q$, that is, $|Q'| = 3$, and then replace every vertex in that class by q copies with the same neighbors in each color in P . If G'_1 and G'_2 are non-isomorphic then the two final graphs are non-isomorphic. Recall that we assume that there are at least three edge colors. Since there are $pk \cdot qk = 3 \cdot pqk$ edges in the complete bipartite graph and since there must be at least $pqk = \text{lcm}(pk, qk)$ edges in a color, there are exactly three colors, say red, green and blue, each with pqk edges. Let $Q' = \{w_0, w_1, w_2\}$. Consider the following two graphs G'_1 and G'_2 . We divide P into three blocks P_0, P_1, P_2 of equal size. We color all edges between P_i and w_i red, edges between P_i and $w_{(i+1) \bmod 3}$ green and the other edges blue, as to form a blown-up color-regular bipartite $K_{3,3}$. This yields the graph G'_1 . To obtain G'_2 we choose for all $i \in \{0, 1, 2\}$ a vertex v_i from P_i and recolor its incident red edge green and its incident green edge red. That is, the edges $\{v_i, w_i\}$ are now green while the edges $\{v_i, w_{(i+1) \bmod 3}\}$ are now red. The newly obtained graph G'_2 is non-isomorphic to G'_1 since in G'_2 there are vertices that agree on some edge colors towards vertices in Q' but disagree on others. (For example, the edge $\{v_0, w_0\}$ is green just like the edges between vertices in $P_2 \setminus \{v_2\}$ and w_0 , but $\{v_0, w_1\}$ is red, whereas P_2 only has blue edges to w_1 . Note that $P_2 \setminus \{v_2\}$ is non-empty since $p > 1$) In G'_1 , for any vertex $v \in P$, the color of any of its edges to Q' uniquely determines to which block of P it belongs and thus also determines the colors of the other edges incident to v . \square

In the lemma, the bipartition classes P and Q are not required to be distinguishable. However, if P and Q are a priori distinguishable (which is, for example, the case if vertices in P are colored with a different color than vertices in Q), then the lemma also applies. The reason why the graphs G'_1 and G'_2 constructed in the proof are still C^2 -

equivalent to G in this situation, is that in case P and Q can be interchanged (which implies $|P| = |Q|$), the color degrees for P and Q are the same. We obtain the following corollary.

Corollary 22. *Let G be an undirected complete bipartite vertex-/edge-colored color-biregular graph on bipartition (P, Q) . Suppose G has at least three edge colors, the vertices in P are colored red and the vertices in Q are colored blue. Then there is a bipartite graph G' on (P, Q) non-isomorphic to G and C^2 -equivalent to G if and only if G is not isomorphic to $K_{3,3}$.*

In the following, we will need both Lemma 21 and Corollary 22 depending on whether we are in a situation where the bipartition classes P and Q can be interchanged. We first use Lemma 21 to characterize color-regular graphs that are identified by C^2 .

Lemma 23. *Let G be an undirected color-regular complete graph with at least three edge colors. Then C^2 identifies G if and only if it is*

1. *a graph on 4 vertices, which has 3 edge colors that each induce a perfect matching,*
2. *a graph on 6 vertices, which has 5 edge colors that each induce a perfect matching,*
or
3. *a graph on 6 vertices, which has 3 edge colors, one of which induces the complement of a 6-cycle, and the other two edge colors each induce an undirected perfect matching.*

Proof: Let $1 \leq d_1 \leq d_2 \leq \dots \leq d_t$ be the color degrees of the graph. Since we interpret non-edges as edges of a particular color, without loss of generality we can assume that $\sum_{i=1}^t d_i = n - 1$.

We distinguish two cases according to the parity of the number n of vertices.

Case 1: n is odd.

Suppose first that n is odd. We will employ the generalized circulant construction outlined in Subsection 3.2. In particular, we saw there that for every map $\psi: \{1, \dots, n-1\} \rightarrow \{1, \dots, t\}$ such that for all i we have $\psi(n-i) = \psi(i)$ and for all j we have $|\psi^{-1}(j)| = d_j$, we obtain a circulant graph $\text{Circ}(\psi)$ by coloring the edge between vertices i and j satisfying $j \leq i$ with the color $\psi(i-j)$.

Since n is odd, all color degrees are even, and since there are at least three colors, we have $1 < d_1 \leq (n-1)/3$. This also implies $n \geq 7$.

There are suitable ψ, ψ' such that $\psi^{-1}(1) = \{-d_1/2, \dots, -1, 1, \dots, d_1/2\}$ and $\psi'^{-1}(1) = \{-d_1/2 - 1, \dots, -2, 2, \dots, d_1/2 + 1\}$. We claim that $\text{Circ}(\psi)$ and $\text{Circ}(\psi')$ are not isomorphic. Indeed, consider vertex 0. In $\text{Circ}(\psi)$ this vertex has two neighbors that both have $d_1 - 2$ common neighbors with it (namely the neighbors 1 and $n-1$). However, since $d_1 \leq (n-1)/3$, in $\text{Circ}(\psi')$ the vertex 0 does not have such a neighbor. Since the graph is transitive, it suffices to consider only vertex 0, and we conclude that the graphs are non-isomorphic.

Case 2: n is even.

Suppose now that n is even. Note first that all color-regular graphs on at most four vertices with at least three edge colors have exactly four vertices and are unions of three perfect matchings. We can thus assume in the following that $n \geq 6$.

For every integer i define $\text{even}(d_i)$ to be the largest even integer that is at most d_i . We distinguish two cases, depending on the following inequality, which we call the *sum condition*. We say that the sum condition is fulfilled if $\sum_{i=2}^t \text{even}(d_i) \geq n/2 - 1$. Note that this summation starts at index 2.

Case 2a: The sum condition is fulfilled.

Observe that if the sum condition is fulfilled then there exists a graph G_H on n vertices indistinguishable from G with the following properties: The vertex set V of G_H is the union of two sets V_1 and V_2 , each of size $n/2$. The two graphs induced by V_1 and V_2 are isomorphic, color-regular graphs and the bipartite graph H induced by all edges that run between V_1 and V_2 is color-regular and contains all edges of color 1. This is possible by applying König's Theorem, since the summation in the sum condition starts at index 2. We will, depending on the color degrees, construct such a graph H as to allow us to also construct a different graph H' that can replace H to obtain a non-isomorphic graph.

($d_1 > 1$.) If $d_1 > 1$ then the graph H can be constructed in such a way that the graph induced by the edges in color d_1 is connected (which can be seen by starting with a Hamiltonian cycle and then applying König's Theorem as in the previous proof). There are at least three colors and d_1 is the smallest color degree, so we have $d_1 \leq (n-1)/3$. Since $n \geq 6$, this implies that $d_1 < n/2 - 1$ and thus, the graph induced by the edges of color 1 is neither a matching nor a co-matching. Lemma 14 implies that there is a bipartite graph H' that is C^2 -equivalent but not isomorphic to H . Let $G_{H'}$ be the graph obtained from G_H by replacing H with H' . Since the edges of color 1 in H induce a connected subgraph, the partition of G_H into the sets V_1 and V_2 is combinatorially determined by the graph G_H (i.e., the partition is isomorphism invariant). If this bipartition is not combinatorially determined by H' then the graphs G_H and $G_{H'}$ are not isomorphic. However, if the bipartition is combinatorially determined by H' then G_H and $G_{H'}$ are also not isomorphic since H and H' are not isomorphic. This resolves the case of the fulfilled sum condition if $d_1 > 1$.

($d_1 = d_2 = 1$.) If $d_1 = d_2 = 1$, we can construct G_H such that all edges of color 1 and 2 are contained in H . We repeat the argument above, this time constructing a bipartite graph H for which the subgraph induced by the edges of color 1 and the edges of color 2 is connected. According to Lemma 21 there is exactly one size for which there exists no non-isomorphic graph H' that is C^2 -equivalent to H , namely the case $n = 6$. Note that in this case, since the sum condition is fulfilled, we have $d_3 = 3$ (which implies $t = 3$), or it holds that $d_3 = 1$ and $d_4 = 2$ (and $t = 4$). In the first case, since two disjoint matchings always form a union of cycles of even length, for every graph with these degrees, the graph induced by the third color must be the complement of a 6-cycle and thus, the graph G is identified by C^2 . In the second case, two non-isomorphic C^2 -equivalent graphs are obtained by forming three matchings and two 3-cycles and by forming three matchings and a 6-cycle.

($d_1 = 1$ and $d_2 > 1$.) Supposing now that $d_1 = 1$ and $d_2 > 1$, consider first the case $n = 6$,

which implies $d_2 = d_3 = 2$. Such a graph is not identified since it is possible to let color d_2 induce two triangles but also possible to let it induce a 6-cycle. Supposing $n > 6$ one can choose the graphs induced by V_1 and V_2 so as to have at least two edge colors. Indeed, since there are at least three edge colors and $d_3 \geq 2$, we can take a number d'_2 with $1 \leq d'_2 \leq d_2$ and let the induced color-regular graphs on V_1 and V_2 be at least d'_2 -regular in color 2 and at least 2-regular in color 3, for example, by using the circulant construction from Subsection 3.1. The graph H contains a perfect matching of color 1. Note that this does not necessarily mean that the bipartition (V_1, V_2) is combinatorially determined by the graph. Since the induced graphs on V_1 and V_2 are isomorphic, the matching can be chosen such that whenever (v_1, v_2) and (v'_1, v'_2) are distinct edges from V_1 to V_2 of color 1 then the color of (v_1, v'_1) is equal to the color of (v_2, v'_2) . However, since $G_H[V_2]$ has at least two edge colors, by suitably changing the matching, we can destroy this property maintaining the C^2 -equivalence class.

We claim that this alteration produces a non-isomorphic graph. We argue this by showing that the number of 4-tuples (v_1, v_2, v'_1, v'_2) of distinct vertices such that (v_1, v_2) and (v'_1, v'_2) are matching edges and (v_1, v'_1) has a different color than (v_2, v'_2) changes. Since the bipartition might not be isomorphism invariant, in addition to the edges considered above, we also need to consider for an unordered pair of matching edges $(\{v_1, v_2\}, \{v'_1, v'_2\})$ with $v_i, v'_i \in V_i$ whether (v_1, v'_2) and (v'_1, v_2) are of the same color (note the reversal of the indices). However, (v_1, v'_2) and (v'_1, v_2) are edges in the graph H . Thus, the number of such pairs for which we obtain edges of different colors does not change if the matching is rearranged while maintaining the isomorphism type of H . We can thus create two C^2 -equivalent graphs which differ in the number of pairs of (directed) matching edges of color 1 whose starting vertices are connected by the same color as the end vertices. These graphs cannot be isomorphic.

Case 2b: The sum condition is not fulfilled.

Suppose now that the sum condition is not fulfilled. Observe that this implies that $d_1 = d_2 = 1$, since otherwise $\text{even}(d_2) + \text{even}(d_3) \geq (d_1 + d_2 + d_3 + 1)/2 - 1$ and $\text{even}(d_i) \geq d_i/2$ for all $i \geq 3$. If there are at most four colors, then $\sum_{i=2}^t \text{even}(d_i) \geq n - 4$. Hence, if $n \geq 6$, the sum condition is fulfilled.

We can thus assume that there are at least 5 colors and that $d_1 = d_2 = 1$. Every map $\psi: \{1, \dots, n-1\} \rightarrow \{1, \dots, t\}$ leads to an edge-colored graph $\text{Match}(\psi)$ via the matching construction from Subsection 3.2. Note that if ψ_1 and ψ_2 are two distinct maps then the graphs $\text{Match}(\psi_1)$ and $\text{Match}(\psi_2)$ are not identical (but may be isomorphic).

We will in the following consider only maps ψ that satisfy $\psi(1) = 1$ and $\psi(2) = 2$.

There are at least $\binom{n-3}{d_3} \binom{n-3-d_3}{d_4} \geq (n-3) \cdot 2$ distinct valid choices for ψ that are of interest to us. Since there are only n possible automorphisms of a color-alternating Hamiltonian cycle on n vertices, if all graphs resulting from the choices for ψ were isomorphic then we would have $(n-3) \cdot 2 \leq n$, which implies $n \leq 6$. We already considered $n \leq 4$ so suppose $n = 6$. In this case, since there must be 5 colors, $d_1 = d_2 = d_3 = d_4 = d_5 = 1$. We thus obtain a 1-factorization of K_6 , of which there is only one up to isomorphism [18], so the graph is identified by C^2 . \square

So far we have only dealt with undirected graphs. We now extend the previous lemma to color-regular complete ec-POGs (see Figure 7). Recall that ec-POGs have the property that there is no edge color with both directed and undirected edges. (With this assumption we do not lose generality since directed edges and undirected edges are always distinguishable.)

Theorem 24. *Let G be a color-regular complete ec-POG. Then C^2 identifies G if and only if G is*

1. *an undirected complete graph with only one edge color,*
2. *undirected and has two edge colors, one of which induces a perfect matching,*
3. *a directed 3-cycle,*
4. *an undirected graph on four vertices, which has three edge colors that each induce a perfect matching,*
5. *a graph on four vertices, has two edge colors, one of which induces a 4-cycle that may or may not be directed, and the other edge color induces an undirected perfect matching,*
6. *the unique regular tournament on five vertices,*
7. *a graph on five vertices, has two edge colors which both induce a 5-cycle of which at most one is directed,*
8. *an undirected graph on six vertices, which has five edge colors that each induce a perfect matching, or*
9. *an undirected graph on six vertices, which has three edge colors, one of which induces the complement of a 6-cycle, and the other two edge colors each induce an undirected perfect matching.*

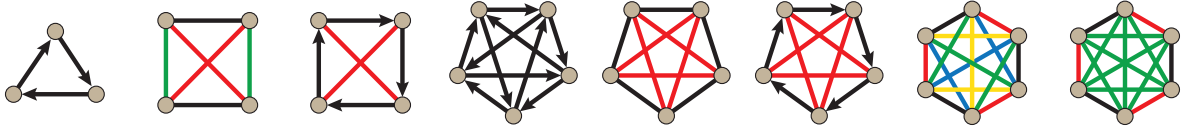


Figure 7: The special cases (Items 3–9) that occur in the classification in Theorem 24.

Proof: Let G be an edge-colored color-regular complete graph.

Let \overline{G} be the graph that is obtained from G by replacing every directed edge with an undirected edge of the same color. We claim that if G is identified by C^2 , then so is \overline{G} . Indeed, assume there is a non-isomorphic graph \overline{H} not distinguished from \overline{G} by C^2 . Then we obtain a non-isomorphic graph H not distinguished from G by directing all edges in edge color classes that consist of directed edges in G . This is consistently possible (i.e., making the graph H color-regular) because each such edge color class induces a regular graph of even degree. By the classical result of Euler we can choose a Eulerian tour, which gives us equal in- and out-degrees for all vertices.

To prove the theorem it thus remains to consider graphs whose undirected underlying graph is identified by C^2 .

(at most 2 edge colors) Starting with graphs that have at most two edge colors we consider all undirected graphs that appear in Lemma 13 (with non-edges interpreted as edges of a distinct color). For graphs with only one edge color, the question is for which n there exists exactly one regular tournament on n vertices. This is the case for $n \in \{1, 3, 5\}$. Indeed, to see that for $n > 5$ there is no unique regular tournament it suffices to consider the circulant directed graph in which there is an edge from i to j if $(i - j) \bmod n < n/2$. Reversing all edges of the form $(i, i + 1)$ yields a non-isomorphic regular tournament.

Next we consider graphs that consist of a matching and the complement of a matching. Suppose the matching is of color red and the complement is blue. Since a matching induces a graph of odd degree, it cannot be directed while maintaining regularity. If $n = 4$ then by directing the complement of the matching we obtain the graph described in Item 5. Note that n is even, so suppose $n > 5$ from now on. We form two non-isomorphic graphs \widehat{G} and \widehat{G}' on $\{v_1, \dots, v_{n/2}\} \cup \{w_1, \dots, w_{n/2}\}$ that are C^2 -equivalent to G as follows. We start with an arbitrary directed complete graph H on vertex set $\{1, \dots, n/2\}$ that contains the edges $(1, 2)$, $(2, 3)$ and $(3, 1)$. In \widehat{G} each pair $\{v_i, w_i\}$ forms an edge of the red matching. For every directed edge (i, j) in H , we insert the blue directed edges (v_i, v_j) , (v_j, w_i) , (w_i, w_j) and (w_j, v_i) . This yields a graph \widehat{G} which consists of a red matching and a blue directed complement of a matching. Note that in \widehat{G} every 4-tuple of vertices that induce two red matching edges also induces a directed 4-cycle in blue. Also note that \widehat{G} contains a blue directed 3-cycle C_3 with vertices v_1, v_2, v_3 . We form a new graph \widehat{G}' by reversing all edges in C_3 . The graph \widehat{G}' is C^2 -equivalent to \widehat{G} , but it is not isomorphic to \widehat{G} since the 4-tuple of vertices (v_1, v_2, w_1, w_2) induces two red matching edges but no blue directed 4-cycle in \widehat{G}' .

Consider now a 5-cycle. Interpreted as an edge-colored complete graph, it has two edge colors, both of which induce a 5-cycle. If we direct one of the 5-cycles, we obtain a graph described in Item 7. It is easy to see that these graphs are identified by C^2 since reversing all edges of the 5-cycle yields an isomorphic graph. However, directing both 5-cycles, i.e., directing edges of both edge colors, we obtain a graph that is not identified by C^2 since it is not isomorphic to the graph obtained when reversing the directions in one of the 5-cycles, since only one of the graphs has a directed 3-cycle with two edges in the first color.

(at least three edge colors) Concerning graphs with three or more edge colors we need to consider all graphs mentioned in Lemma 23. Since matchings cannot be oriented maintaining regularity it suffices to consider graphs on 6 vertices that are the disjoint union of the complement of a 6-cycle and two perfect matchings. However, for the edge color class that is the complement of a 6-cycle the degree is odd, so it cannot be oriented in a regular way either. We thus obtain only the undirected versions of the graphs mentioned in Lemma 23. \square

We are going to determine the structure of ec-POGs that are identified by C^2 . So far, we have classified the ec-POGs that have exactly one C^2 -partition class. Similarly to the case of undirected graphs treated in Section 4 we will now describe how to combine

such building blocks to form a larger graph that is identified. As before, by interpreting non-edges as edges of a special color we only need to consider complete graphs. Since vertices of different C^2 -partition classes are distinguished, we can assume that all edges between two C^2 -partition classes are undirected. (In other words, the direction of a directed edge starting in a color class P and ending in another color class Q is implied by the colors of the starting and end vertices. Edges of the same color in the other direction can thus always be marked with a different color). Recall that Corollary 22 says that the only regular graph with a fixed bipartition and at least three edge colors that is color-biregular is $K_{3,3}$ with an edge coloring that induces three disjoint perfect matchings. This shows that it is possible that two C^2 -partition classes of size 3 in an identified graph can be connected via three matchings of different colors. In the light of this, in comparison to the case of two colors covered in Lemma 14 we need an additional relation.

Definition 25. Let G be a vertex-colored ec-POG and let P and Q be two disjoint subsets of $V(G)$. We introduce the relation $P \equiv_3^3 Q$ to denote the fact that the graph induced by the edges running between P and Q is the graph $K_{3,3}$ with three edge colors which each induce a perfect matching between P and Q .

Since we treat non-edges as edges of a particular color, we also need to adapt our notation which we used for graphs (see Notation 15) so that it becomes applicable to graphs with two edge colors. Suppose between two distinct C^2 -partition classes P and Q there are edges of two colors, say red and blue. To define the relations $P \sqcap Q$, $P \doteq Q$, $P \ll Q$ we always consider the graph induced by the edge color class that contains fewer edges and ignore orientations. It is not difficult to see that if the number of red edges is equal to the number of blue edges then choosing either induced graph yields the same results. Note that with this convention the relations $P \sqcap Q$, $P \doteq Q$, $P \ll Q$ in particular imply that there are only at most two colors among the edges running between P and Q .

Corollary 26. Let G be a vertex/edge-colored undirected graph that is identified by C^2 . If P and Q are distinct C^2 -partition classes of G , then $P \sqcap Q$, $P \doteq Q$, $P \ll Q$, $Q \ll P$ or $P \equiv_3^3 Q$.

Theorem 24 implies that each C^2 -partition class of an ec-POG that is identified must be one of 9 listed types. In an identified ec-POG, we call every C^2 -partition class which does not induce an undirected complete graph with only one edge color an **exception** (i.e., any class that does not fall under Item 1 of Theorem 24). Similarly, we call every pair of C^2 -partition classes P and Q for which $P \equiv_3^3 Q$ holds an *exception*.

Note that with this terminology Lemma 16 says that for uncolored, undirected graphs each connected component of the skeleton can induce at most one exception in the original graph. Here, in accordance with Lemma 13, the possible exceptions are Item 2 and the undirected option in Item 7 from Theorem 24. As we will see, a similar observation about exceptions only occurring once per connected component of the skeleton is true for edge-colored graphs.

Since in an edge-colored graph it is not clear what a flip is, we need to slightly adjust the approach from Section 4. Let G be a vertex/edge-colored graph. As before, we define the vertices of the *skeleton* S_G to be the C^2 -partition classes of G . Two distinct vertices P, Q in S_G are adjacent in S_G if the corresponding classes in G do not satisfy $P \sqcap Q$, i.e.,

whenever P and Q are not monochromatically connected.

Concerning the properties of identified structures, we obtain a theorem similar to Corollary 20 for graphs.

Theorem 27. *Let G be a vertex-colored ec-POG (i.e., a vertex/edge-colored partially oriented graph). Then G is identified by C^2 if and only if the following hold:*

1. *Each C^2 -partition class induces a graph identified by C^2 (i.e., the induced graph is one of the graphs mentioned in Theorem 24),*
2. *for all C^2 -partition classes P and Q we have $P \sqsubset Q$, $P \dot{=} Q$, $P \ll Q$, $Q \ll P$ or $P \equiv_3^3 Q$,*
3. *the skeleton S_G is a forest,*
4. *there is no path P_0, P_1, \dots, P_t in S_G with $P_0 \ll P_1$ and $P_{t-1} \gg P_t$,*
5. *there is no path P_0, P_1, \dots, P_t in S_G where $P_0 \ll P_1$ and P_t is an exception, and*
6. *in every connected component of S_G there is at most one exception (i.e., either two C^2 -partition classes P and Q satisfying $P \equiv_3^3 Q$ or a C^2 -partition class among Items 2–9 described in Theorem 24).*

Proof: For the backward direction let G first be a vertex-colored ec-POG that satisfies the given conditions. It suffices to show that for each connected component in the skeleton S_G , the graph induced by its collection of C^2 -partition classes is identified by C^2 . For such an induced graph note that, if the entire graph itself is not already an exception, then it contains a non-exceptional C^2 -partition class P such that for exactly one class $Q \neq P$ we have $P \ll Q$ or $P \dot{=} Q$ and for all other C^2 -partition classes $Q' \neq Q$ (including P itself) we have $P \sqsubset Q'$. Such a graph is identified if and only if the graph obtained by removing P is identified. It thus suffices to show that all exceptions are identified. This is shown in Corollary 22 and Theorem 24.

For the forward direction let G be a vertex-colored ec-POG identified by C^2 . Note that the vertex coloring of G only has the effect that the C^2 -partition is refined. Corollary 20 and Theorem 24 imply Part 1. Corollary 20 and Corollary 22 imply Part 2.

Note the following for two C^2 -partition classes P and Q of an identified graph that satisfy $P \equiv_3^3 Q$. If we merge two of the three edge color classes into a single one to obtain a graph G' , then the entire graph must still be identified by C^2 . The reason is that if two classes P and Q of size 3 satisfy $P \dot{=} Q$ then the larger edge color class can always be split so that $P \equiv_3^3 Q$. Similarly, if we merge different edge color classes within a C^2 -partition class P so that the graph induced by P is still identified then the entire graph must still be identified by C^2 . Likewise, if we remove directions from some color classes of edges within P so that the graph induced by P is identified then the entire graph must still be identified by C^2 .

In particular, we can replace every C^2 -partition class by an undirected complete graph while maintaining that the graph is identified.

With this in mind, Parts 3–5 follow immediately from Corollary 20.

To prove Part 6 we first ignore the \equiv_3^3 exceptions. Note that by Parts 4 and 5 exceptions of the other types that occur in the same component of S_G must have the same size. Moreover, these exceptions must be connected by a path P_0, P_1, \dots, P_t such that we have $P_0 \dot{=} P_1 \dot{=} \dots \dot{=} P_t$ in G . Suppose in such a situation P_0 and P_t each induce an exception. Note that for all such exceptions apart from the tournaments on 3 and 5 vertices, it is possible to remove directions and merge colors in order to obtain a matching or a 5-cycle. It follows from Corollary 22 that such exceptions cannot occur in the same component of the skeleton. If P_0 and P_t are directed 3-cycles, then by reversing the directions in P_t we obtain a non-isomorphic graph. Similarly, if P_0 is the regular tournament on 5 vertices and P_t is an arbitrary exception on 5 vertices with some directed edges then we obtain a non-isomorphic graph by reversing all edges in P_0 . If P_t is an exception on 5 vertices with no directed edges then P_t must be the exception that has two edge colors which each induce an undirected 5-cycle. Since P_0 contains 5-cycles of the underlying undirected graph for which the edges form a directed cycle and 5-cycles for which the edges do not form a directed 5-cycle, by permuting the vertices in P_t we can construct a non-isomorphic graph (by either choosing to match a 5-cycle of P_t with a directed 5-cycle of P_0 or with a 5-cycle whose edges do not form a consistent cyclic orientation).

Suppose finally that an exception of the type \equiv_3^3 occurs in a connected component of S_G . Then by merging colors and by Parts 4 and 5 the only types of exceptions that can occur in the connected component are other exceptions of type \equiv_3^3 and directed 3-cycles. Observe that an exception of type $P \equiv_3^3 Q$ induces a cyclic order on both P and Q . Thus, by swapping two edge colors of the \equiv_3^3 -exception we obtain a non-isomorphic graph whenever there is a second exception. \square

Corollary 28. *Given a vertex-colored ec-POG on n vertices we can decide whether it is identified by C^2 in time $O(n^2 \log n)$.*

Proof: For this we compute the coarsest equitable partition in time $O(n^2 \log n)$ using the color refinement procedure. We then compute the skeleton of the obtained graph. For each C^2 -partition class we check whether the induced graph appears among the exceptions from Theorem 24. Since the non-trivial exceptions have at most 6 vertices, this can be performed in linear time. We also check whether each pair of distinct C^2 -partition classes satisfies one of the relations required by Corollary 26. We mark each vertex or edge of the skeleton that corresponds to an exception. Finally, we check whether the skeleton is a tree, whether each connected component has at most one exception, and whether the sizes of the C^2 -partition classes in each connected component monotonically increase when starting from a smallest class (for example using depth first search). We also check that every exception is a smallest class of its connected component. \square

Since we can reduce the problem of deciding whether a finite relational structure is identified by C^2 to the problem of deciding whether an ec-POG is identified by C^2 we obtain the following corollary.

Corollary 29. *Given a finite relational structure \mathfrak{A} with a universe of size n over a fixed signature we can decide in time $O(n^2 \log n)$ whether it is identified by C^2 .*

Proof: Since the logic C^2 can only use two variables, if $n \geq 3$ and \mathfrak{A} has a relation of arity at least 3, then it is not identified. For a fixed signature and a given structure \mathfrak{A}

over this signature, the graph $\text{ec-POG}(\mathfrak{A})$ can be constructed in $O(n^2)$ time. It suffices now to observe that a relational structure \mathfrak{A} with relations of arity at most 2 is identified if and only if $\text{ec-POG}(\mathfrak{A})$ is identified. \square

Using the classification we also obtain an extension of Corollary 19 to ec-POGs and, more generally, to finite relational structures.

Corollary 30. *Let G be an ec-POG and χ a vertex coloring of G such that (G, χ) is identified by C^2 and let χ' be a vertex coloring of G which induces a finer partition on $V(G)$ than χ does. Then (G, χ') is also identified by C^2 .*

Proof: By induction, it suffices to consider a vertex coloring that refines exactly one C^2 -partition class of G . First note that every vertex-colored version of an undirected complete graph has no exception. Furthermore, for a graph that is a perfect matching, if we consider the C^2 -partition classes obtained after coloring the vertices arbitrarily, then each component of the skeleton induces at most one exception (which must be a matching).

Next note that every vertex-colored version of a graph described in Items 3–9 of Theorem 24 is identified and has at most one exception. This is easy to see for all described graphs up to 5 vertices. For the mentioned graphs on 6 vertices consider first the case of 5 disjoint matchings. It is not difficult to see that this graph only has two equitable partitions, the discrete partition and the trivial partition. Consider now the graph on 6 vertices that consists of two perfect matchings and a 3-regular graph. Every equitable partition with a singleton for this graph is discrete. There are thus two types of equitable partitions to consider, namely equitable partitions with three classes of size 2 and equitable partitions with two classes of size 3. In the latter case, the graph induces an edge-colored version of $K_{3,3}$ and in the former case the graph has at least two classes P_1 and P_2 such that $P_1 \square P_2$ (since each of the two matchings can only appear in one biregular connection between two classes). In any case, the graph is identified and has at most one exception. This proves that Properties 1 and 6 of Theorem 27 are preserved.

To show globally that a color-refined version of an identified ec-POG is also identified, note that also all other properties required by Theorem 27 are preserved, since (by Corollaries 19 and 20) these properties are preserved if all C^2 -partition classes are replaced by undirected complete graphs. \square

Corollary 31. *If a finite relational structure \mathfrak{A} is identified by C^2 , then every finite relational structure obtained from \mathfrak{A} by adding unary relations is also identified by C^2 .*

Proof: Since \mathfrak{A} is identified, it has at most 2 elements in its universe or there is no relation of arity at least 3. Recall that under this circumstance, \mathfrak{A} is identified if and only if $\text{ec-POG}(\mathfrak{A})$ is identified. The addition of unary relations corresponds to a refinement of the vertex coloring of $\text{ec-POG}(\mathfrak{A})$, so the corollary follows from the previous one. \square

6 Higher dimensions

One may wonder whether the results presented for C^2 throughout this paper can be extended to the logics C^k . Chang [7] and Hoffman [14] showed that triangular graphs on

sufficiently many vertices are identified by C^3 . They form an infinite family of strongly regular graphs, which shows that for $k > 2$ any classification result for graphs identified by C^k must include an infinite number of non-trivial graphs. This already indicates that the situation for $k = 2$ is special. We first show that statements analogous to Corollaries 6 and 19 do not hold for higher dimensions. We use the construction from [6] which provides us with the following fact.

Fact 32 (Cai, Fürer, Immerman [6]). *For every $k > 2$, there are non-isomorphic 3-regular graphs G and G' of size $O(k)$ on the same vertex set V , which contains vertices a, b, a', b' with the following properties: The graphs are not distinguishable by C^k and the edge sets satisfy $E(G) \setminus E(G') = \{\{a, a'\}, \{b, b'\}\}$ and $E(G') \setminus E(G) = \{\{a, b'\}, \{b, a'\}\}$. Moreover, we can guarantee that every graph indistinguishable from G and G' by C^3 is isomorphic to G or G' .*

The graphs G and G' referred to in the fact are obtained by replacing every vertex v of a suitable connected graph by a gadget that only depends on the degree of v (see [6] for details). Each original edge $\{v, v'\}$ corresponds to two disjoint edges in the new graph, connecting a pair of specified vertices $\{a_v, b_v\}$ in the gadget for v to a pair of specified vertices $\{a_{v'}, b_{v'}\}$ in the gadget for v' . If the two disjoint edges are $\{a_v, a_{v'}\}$ and $\{b_v, b_{v'}\}$, we call them *parallel*. Otherwise, if they are $\{a_v, b_{v'}\}$ and $\{b_v, a_{v'}\}$, we call them *twisted* and the pair of them a *twist*.

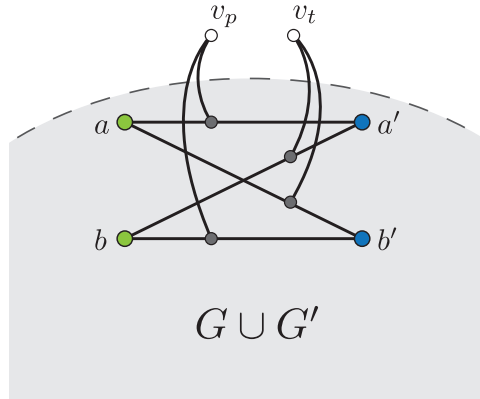


Figure 8: the construction used to show that higher-dimensional analogues of Corollaries 6 and 19 do not hold

Each gadget consists of so-called middle vertices and d pairs of outer vertices, where d is the degree of the corresponding vertex in the original graph. Cai, Fürer and Immerman show the following. When choosing each pair of edges between gadgets of adjacent vertices to be arbitrarily either parallel or twisted, we can only obtain one of two graphs up to isomorphism. More precisely, the isomorphism type of the resulting graph only depends on the parity of the number of twists in it.

In [6], first colored graphs are constructed. By a standard transformation which reduces isomorphism of colored graphs to isomorphism of uncolored graphs, one can remove the colors in the graph. Since there are suitable transformations that do not add more than $O(k)$ extra vertices, we can assume the graphs G and G' appearing in Fact 32 to be

uncolored, still preserving the property that for a fixed input graph, all possible resulting graphs with the same parity of the number of twists are isomorphic. We will continue to work with colored graphs, knowing that the same reduction techniques can be applied.

Every gadget of the construction is, when considered by itself, identified by C^3 . This can easily be verified by using the fact that the pairs of outer vertices have distinct colors and every two of them form a simple 8-cycle together with the middle vertices. It is straightforward to conclude that a graph that is indistinguishable from G or G' may only differ from them by the number of twists it has. Depending on that parity, the graph is either isomorphic to G or to G' .

Theorem 33. *For every $k > 2$, there is a graph H of size $O(k)$ identified by C^3 for which the C^k -partition is strictly coarser than the orbit partition. Moreover, not all vertex-colored versions of H are identified by C^k .*

Proof: Let G and G' be the graphs described in Fact 32. We form $G \cup G'$, which is the graph on $V(G) = V(G')$ with edge set $E(G) \cup E(G')$. This graph in particular has the parallel edges $\{a, a'\}, \{b, b'\}$ and the twisted edges $\{a, b'\}, \{b, a'\}$. Subdivide these four edges and connect the two middle vertices of the subdivided parallel edges to a new vertex v_p and the two middle vertices of the subdivided twisted edges to a new vertex v_t . Let H be the resulting graph.

To argue that H is identified by C^3 , we will use the fact that in the Cai, Fürer, Immerman construction we can achieve that vertices in different gadgets have different colors. (Without this, it can still easily be proven that C^4 identifies the graph.) Since every gadget is identified by itself, the only critical edges are the ones connecting gadgets. They can either be twisted or parallel. However, since $G \cup G'$ has both versions present between the vertices a, a', b, b' , for every inserted twist, there is an isomorphism to $G \cup G'$ that resolves the twist. Thus, the graph is identified since it is isomorphic to its twisted version. Consequently, the graph H is identified by C^3 since the newly added vertices v_p and v_t are the only vertices of degree 2 and therefore distinguished by C^3 from all other vertices.

It holds that v_p and v_t are indistinguishable by C^{k-4} . (This can easily be seen by considering pebble games. Spoiler could win the game on G and G' by simulating the game on H , placing 4 additional pebbles on a, b, a' and b' , which would imply that G and G' can be distinguished by C^k .)

We claim that v_p and v_t are not in the same orbit. Assuming otherwise, there is an automorphism that maps the neighbors of v_p to the neighbors of v_t and vice versa. This implies that there is an automorphism that maps the parallel edges to the twisted edges and vice versa. However, this implies that G and G' are isomorphic, yielding a contradiction.

Note that, since v_p and v_t are indistinguishable by C^{k-4} , the graph obtained by coloring v_p with a color different from the color of every other vertex is not identified by C^{k-4} . \square

Even if a graph is identified and the orbits of its automorphism group are correctly determined by C^k , it may still be the case that this does not hold for all colored versions of the graph.

Theorem 34. *For every $k > 2$, there is a graph H of size $O(k)$ which is identified by C^3 such that the C^k -partition classes are the orbits of H but there are vertex-colored versions of H that are not identified by C^k and for which the C^k -partition classes are not the orbits of H .*

Proof: We slightly modify the construction outlined in the previous proof. We first subdivide the parallel edges $\{a, a'\}$ and $\{b, b'\}$ to obtain undirected paths $(a = a_1, a_2, a_3 = a')$ and $(b = b_1, b_2, b_3 = b')$, respectively, and also insert all twisted edges $\{a_i, b_{i+1}\}$ and $\{b_i, a_{i+1}\}$. As before, we subdivide all new edges and add for $i \in \{1, 2\}$ vertices $v_{i,p}$ adjacent to the midpoint of $\{a_i, a_{i+1}\}$ and to the midpoint of $\{b_i, b_{i+1}\}$. We also add vertices $v_{i,t}$ adjacent to the midpoint of $\{a_i, b_{i+1}\}$ and to the midpoint of $\{b_i, a_{i+1}\}$. We obtain a graph H that is, by the same reasoning as above, identified by C^3 . Furthermore, the logic C^3 determines the orbits of H since for $i \in \{1, 2\}$ the vertex $v_{i,t}$ can be mapped to $v_{i,p}$ by an automorphism. Coloring $v_{1,p}$ yields a graph whose orbits are not the C^k -classes since $v_{2,p}$ and $v_{2,t}$ are not in the same orbit but not distinguished by C^k , i.e., they are in the same C^k -partition class. \square

By repeating the subdivision step, the construction described in the proof can easily be generalized to show that there exist vertex-colored graphs of size $O(k)$ in which orbits are correctly determined by C^3 even if k vertices are individualized, but in which there are $k + 1$ vertices such that individualization of all of them produces a graph whose C^k -partition classes are not the orbits of the automorphism group.

Acknowledgments

We thank Konstantinos Stavropoulos and Martin Grohe for providing us with inspiring, helpful comments that guided us towards the inversion problem.

References

- [1] V. Arvind, J. Köbler, G. Rattan, and O. Verbitsky. Graph Isomorphism, Color Refinement, and Compactness. *ArXiv e-prints*, Feb. 2015.
- [2] A. Atserias and E. N. Maneva. Sherali-adams relaxations and indistinguishability in counting logics. *SIAM J. Comput.*, 42(1):112–137, 2013.
- [3] L. Babai, P. Erdős, and S. M. Selkow. Random graph isomorphism. *SIAM J. Comput.*, 9(3):628–635, 1980.
- [4] L. Babai and L. Kucera. Canonical labelling of graphs in linear average time. In *FOCS'79*, pages 39–46. IEEE Computer Society, 1979.
- [5] C. Berkholz, P. S. Bonsma, and M. Grohe. Tight lower and upper bounds for the complexity of canonical colour refinement. In H. L. Bodlaender and G. F. Italiano, editors, *ESA 2013*, volume 8125 of *LNCS*, pages 145–156. Springer, 2013.

- [6] J. Cai, M. Fürer, and N. Immerman. An optimal lower bound on the number of variables for graph identifications. *Combinatorica*, 12(4):389–410, 1992.
- [7] L.-C. Chang. The uniqueness and nonuniqueness of the triangular association scheme. *Sci. Record.*, 3:604–613, 1959.
- [8] A. Dawar and B. Holm. Pebble games with algebraic rules. In *ICALP 2012*, volume 7392 of *LNCS*, pages 251–262. Springer, 2012.
- [9] E. Grädel and M. Otto. On logics with two variables. *Theor. Comput. Sci.*, 224(1-2):73–113, 1999.
- [10] E. Grädel, M. Otto, and E. Rosen. Two-variable logic with counting is decidable. In *LICS’97*, pages 306–317. IEEE Computer Society, 1997.
- [11] M. Grohe. Finite variable logics in descriptive complexity theory. *Bull. Symbolic Logic*, 4(4):345–398, 1998.
- [12] M. Grohe and M. Otto. Pebble games and linear equations. In P. Cégielski and A. Durand, editors, *CSL’12*, volume 16 of *LIPICs*, pages 289–304, 2012.
- [13] L. Hella. Logical hierarchies in PTIME. *Inf. Comput.*, 129(1):1–19, 1996.
- [14] A. J. Hoffman. On the uniqueness of the triangular association scheme. *Ann. Math. Statist.*, 31(2):492–497, 06 1960.
- [15] E. Kopczynski and T. Tan. Regular graphs and the spectra of two-variable logic with counting. *CoRR*, abs/1304.0829, 2013.
- [16] A. Krebs and O. Verbitsky. Universal covers, color refinement, and two-variable logic with counting quantifiers: Lower bounds for the depth. *CoRR*, abs/1407.3175, 2014.
- [17] L. Kucera. Canonical labeling of regular graphs in linear average time. In *FOCS’87*, pages 271–279. IEEE Computer Society, 1987.
- [18] C. C. Lindner, E. Mendelsohn, and A. Rosa. On the number of 1-factorizations of the complete graph. *J. Comb. Th. Series B*, 20(3):265 – 282, 1976.
- [19] E. Lucas. *Récréations mathématiques*. 2ième éd., nouveau tirage. Librairie Scientifique et Technique Albert Blanchard, Paris, 1960.
- [20] B. D. McKay and A. Piperno. Practical graph isomorphism, II. *J. Symb. Comput.*, 60:94–112, 2014.
- [21] M. Otto. Canonization for two variables and puzzles on the square. *Ann. Pure Appl. Logic*, 85(3):243–282, 1997.
- [22] I. Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *J. Log. Lang. Inf.*, 14(3):369–395, 2005.
- [23] D. B. West. *Introduction to Graph Theory (2nd Edition)*. Pearson, 2000.